Project Overview

Your task is to build a small data pipeline and exploratory data analysis (EDA) on NYC 311 Service Requests from NYC Open Data website. The project will be implemented in Python using Jupyter notebooks, version controlled using Git, and shared via a public GitHub repository. Feel free to ingest other datasets as you see fit, e.g., population of each borough, zip code to borough mapping, etc. Do not use tools like ChatGPT to generate the code, we will follow up with a Zoom to walk through the project.

Dataset: 311 Service Requests from 2010 to Present

1. Set Up Your Project Repository

- Create a clean GitHub repository**.**
- Include a README.md.

2. Data Ingestion Notebook

- Ingest data Python and Pandas.
- Clean and wrangle the data as you see fit.
- Write out the data locally.

3. Exploratory Data Analysis (EDA) Notebook

- Perform basic data profiling, below are some examples, but please add what you feel are useful.
    - Missing value analysis
    - Top complaint types
    - Distribution of requests over time
    - Requests per borough, request per capita for each borough
- Plot at least 3 meaningful visualizations using tools like:
    - matplotlib / plotly
- Extract insights or patterns (e.g., seasonal trends, complaint spikes)

Deliverables

- Jupyter notebooks:
    - 01_data_ingestion.ipynb
    - 02_eda.ipynb
- README.md
- Requirements file: requirements.txt or pyproject.yml
- GitHub repo URL (public or private shared with our team)